

FusedRF: Fusing Multiple Radiance Fields

Rahul Goel¹

Dhawal Sirikonda¹

Rajvi Shah²

PJ Narayanan¹

¹CVIT, KCIS, IIT Hyderabad

²Meta Reality Labs, USA



Figure 1. *NeRF-Feast*: Our method can be used to create a single composed representation. Here, we show a feast prepared on the table from the garden scene from MipNeRF360 [3]. Since we have a single fused representation, the scene is rendered at the cost of a single RF representation while also occupying a memory footprint of a single RF. Scene components are picked from [1]

Abstract

Radiance Fields (RFs) have shown great potential to represent scenes from casually captured discrete views. Compositing parts or whole of multiple captured scenes could greatly interest several XR applications. Prior works can generate new views of such scenes by tracing each scene in parallel. This increases the render times and memory requirements with the number of components. In this work, we provide a method to create a single, compact, fused RF representation for a scene composited using multiple RFs. The fused RF has the same render times and memory utilizations as a single RF. Our method distills information from multiple teacher RFs into a single student RF while also facilitating further manipulations like addition and deletion into the fused representation.

1. Introduction

The use of Radiance Fields (RF) to represent scenes has been of great interest lately. Given a discrete set of posed images of the scene, NeRF [22] utilizes simple MLPs to

encode the scene allowing the synthesis of novel views. While MLPs encoded the scene faithfully, the training time to learn the representation was high, in the range of 8-10 hours. Subsequent efforts like Plenoxels, DVGO [30,31,42] exploited explicit 3D lattice structure to represent the scene and reduced the scene learning times to ~ 10 minutes. Later extensions like TensorRF [6] compressed the volumetric lattice representation while maintaining low learning times.

As the efficiency of scene representation improved, efforts were made towards editing the geometric content of the RFs. While efforts like [23, 36] aimed at removing the existing content and inpaint the missing regions, works like [8, 15, 16, 20, 21, 28, 33, 39, 43, 45] aimed at deforming the RF using cage-based and topological deformations.

Unlike these, we are interested in compositional editing, in which objects and parts from multiple RFs are combined into a single, compact, *fused* RF representation. Compositing for view generation via parallel tracing rays into multiple RF scenes has been attempted before [2, 10, 18, 35, 37, 38, 40, 41]. As multiple component RFs that make up the scene need to be kept simultaneously, the memory footprint and rendering time increase in this

method. Most prior works use the compact neural representation of RFs as a result. On the other hand, ControlNeRF [18] uses full volumetric lattice representation with a common rendering network for all component RFs. This approach demands higher memory, specifically when more than two scenes are to be composited. Additionally, the render times are proportional to the number of RFs that make up the scene due to the need to shoot a large number of rays into each RF.

High memory requirements and rendering times are undesirable for use cases of using the captured RF content in virtual environments, creating extended (XR) and augmented reality (AR) applications. This use requires lower memory and fast rendering of the final composited scene. One way to achieve this goal is to create a single representation by training an RF from scratch using a large number of rendered RGB views. The resulting radiance field will be a single, compact, composited RF with similar memory footprints and render times as a single RF. This approach is, however, inefficient as it involves complete view generation and retraining which could be expensive and slow. Given that all parts in the composited scene already use RF representation, is there a way to *fuse* them directly into an RF representation without expensive RGB retraining?

In this paper, we propose a method to *fuse* multiple RFs into a single one using distillation on their RF [11]. Our distillation is similar to the step used by PVD to translate one type of RF to another [11]. We iteratively fuse the source RFs (with affine composition) into a single, fused RF. Distillation can be performed directly on the σ and color values stored in the RF quickly in a supervised manner for acceptable results. The quality can be improved using a few additional iterations of RGB-based training at the end. The fused RF representation has the same memory requirements and rendering speed as a single RF. Tab. 1 presents the features of our system relative to others that can also do view generation of composited scenes using parallel ray tracing.

2. Related Work

Due to space constraints, we restrict this section to the following relevant works.

Editing Radiance Fields: Several works have been proposed to edit Radiance Fields. The works like [4, 5, 26, 29, 44] aim to edit the appearance by manipulating the light transport equation, works like [7, 14, 27] utilize image-based priors for the stylization of radiance fields. On the other hand works like [8, 15, 20, 39] work on the deformation of radiance fields, and few works concentrate on compositing and creating a novel scene from multiple individually captured scenes [18, 35, 38, 40, 41]. We aim to create novel scenes by compositing captured scenes that have applications in XR.

Segmentation of Scene Content: Often composition of

Works	Native		Composition	
	Small Memory Footprint	Short Training Time	Small Memory Footprint	Short Render Times
NeRF [22]	✓	✗	✓	✗
D ² NeRF [38]	✓	✗	✓	✗
PlenOxels [12]	✗	✓	✗	✗
DVGO [30]	✗	✓	✗	✗
InstantNGP [25]	✗	✓✓	✗	✓✓
CNeRF [18]	✗	✓	✗	✗
TensoRF [6]	✓	✓	✓	✗
PVD [11]	✓	✓	✗	✗
FusedRF (Ours)	✓	✓	✓	✓

Table 1. While works like NeRF and D²NeRF struggle in render times, works like DVGO, Plenoxel, and ControlNeRF additionally also demand high memory. Leading to the infeasibility of the composition of more than a few scenes. Our method efficiently fuses the RFs and maintains memory, and renders times to a single RF

these scenes involves partial addition of semantically correct content captured in one scene to another. Obtaining such semantically correct partial scene content is tackled in recent works either by fusing semantic features [13, 17, 34] into RFs or propagating multi-view 2D masks or labels into the radiance field representations [24, 46]. For this task, we choose [13], which is known to present the most accurate segmentation (extraction) of radiance fields.

Fusion of multiple Radiance Fields: A simple composition using affine transformation has been attempted by works like D²NeRF [38] where compact Neural-RFs were employed to obtain the desired composition. The desired composition increases memory footprints and renders times in proportion to the number of scenes involved. In the case of explicit volumetric lattice, though the partial content retrieval is more accurate, the memory requirement grows rapidly, leading to infeasibility when more than a few scenes are being composited. Hence, a fused representation of composition is desired owing to memory footprints and render times as that of a single RF.

Efficient Fusion: Works like PVD [11] provide methodologies for faster distillation of one RF to another in a supervised setting. But the work is limited to a single RF representation. We draw ideas from this work and build a Fuse-dRF representation that is compact and easy to render.

3. Method

Fig. 2 shows an overview of our FusedRF method. The following through Secs. 3.1 to 3.3 will detail the compositional, fusion, and convergence aspects of the proposed methodology, respectively.

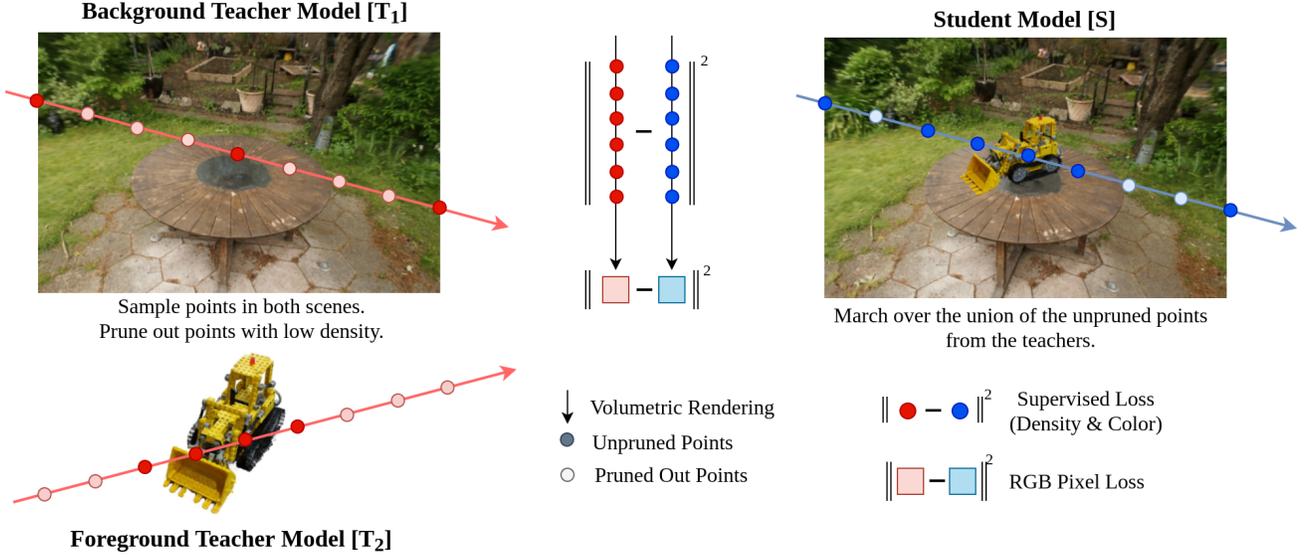


Figure 2. *FusedRF*: Method to fuse multiple RFs into single presentation. We shot the same ray into both teachers T_1 , T_2 (left) created using ISRF [13] to distill a combined Student S (right). For every sampled point on a ray of both RFs we prune out ones with lower densities and apply supervised distillation losses to obtain faster convergence. Followed by a few iterations of Pixel Loss on alpha-composited RGB for smoothening. By fusing multiple RFs into a single representation, we reduce rendering and memory overheads of composition. Both scenes GARDEN and LEGO are picked from the real-world [3] dataset.

3.1. Composition of Radiance Fields

The composition of two distinct radiance fields can be performed by altering the compositional aspect of the volumetric rendering equation [22, 30]. For simplicity, let’s assume we have to composite only two radiance fields (RF^1 and RF^2). Every ray is shot and sampled similarly in both RFs. For every sampled point along the ray, the activated volumetric density ($\alpha : \alpha^I = 1 - e^{-\sigma^I \delta^I}$) is calculated, where I corresponds to the respective Radiance Field (RF^I) [30]. The point with a higher α value is chosen for a contribution towards the rendering of color and density. The resultant RF is considered ground truth for scene composition. It can be observed that the tracing and sampling of the same ray twice are redundant; subsequently, it causes high render times and larger memory footprints. To address these issues, we fuse the RFs into a single representation.

3.2. Fusing Radiance Fields

Rendering two or more radiance fields simultaneously for composition is computationally expensive and cannot be scaled as the memory and computation increase linearly with the number of radiance fields involved in composition. To this end, we propose a method that quickly distills from multiple RFs to a single representation. The resulting representation is as compact and efficient as a single RF. We leverage the already learned 3D information to take losses in 3D, which leads to faster learning (distillation).

Let the two radiance fields to be composed be T_1 and T_2

(teacher1 and teacher2), and the final fused radiance field be S (student). For brevity, let us assume we do not apply any rigid transformation on the radiance fields. We shoot a ray through both T_1 and T_2 and sample points on the ray. The points with low density [○] are pruned out while the ones with high density [●] are utilized. The union of these selected points from T_1 and T_2 acts as our training set for the student S [●]. We query the three radiance fields (T_1, T_2, S) for their density, alpha and color (σ, α, c) at every training sample point location. We apply a *Supervised loss* $\| \bullet - \bullet \|_2^2$ to the color and density values of the student (S) against the corresponding teachers (T_1, T_2) at every training point selected above. This supervised distillation will fuse the composition into one single scene.

As a final stage, we render the RGB values for the rays by accumulating the individual color values weighted by the activated volumetric density using Volumetric Rendering Equation [32] and take an alpha-composited *RGB pixel loss* $\| \square - \square \|_2^2$ for a few iterations. This helps smooth the result around the boundaries of the inserted object.

3.3. Fast convergence

To obtain a single representation of a composed radiance field, one could use the traditional *RGB loss* against the rays from composed RFs or rendered views extracted from the composition. But this would essentially be re-training and would amount to the same time as training an RF from scratch. However, since we have 3D information from already-trained RFs, we can leverage the *supervised*

losses employed at every sampled point, which achieves faster convergences. The augmented convergence is due to 3D distillation. Pruning of low-density points suggested in Sec. 3.2 further speeds up the process.

Additionally, it is often the case that during the composition of radiance fields, one of the scenes is in the majority (dubbed as a background scene). Initializing the student with the background scene significantly speeds up the distillation process. Hence, we initialize the student representation S with the weights of the background scene (one of the dominant teachers T_i). The reduction in time in the case of our distillation-based fusion against total retraining is $3\times$.

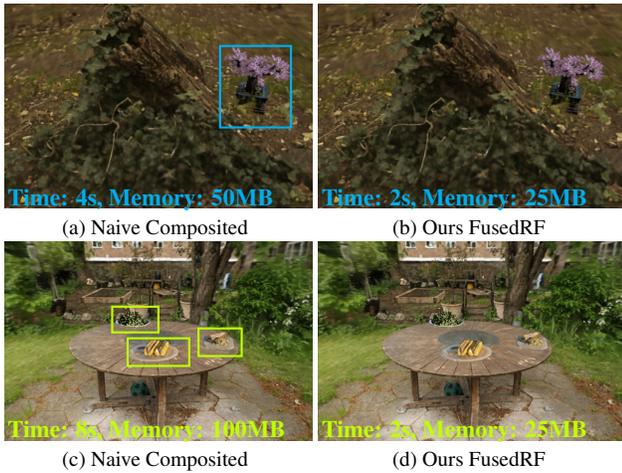


Figure 3. The figure shows results of compositing multiple RF into a single scene, (left column Figs. 3a and 3c) shows results of composition (right column Figs. 3b and 3d) shows results of our FusedRF. Respective memory footprints and rendering times are mentioned in the insets. Figs. 3a and 3b are two scenes picked from [3] and Figs. 3c and 3d is combination of [3] and synthetic scenes.

4. Results

Performance: To validate the performance of our method, we provide Render times and Memory demands against other means of composition, namely 1) Neural-RF, 2) Explicit lattice structures 3) Fused Representation. We tabularize these results in the Tab. 2. It can be observed that the render times and memory footprints increase linearly in the case of Neural-RFs [22] and Explicit lattice representations [6, 12, 30, 31] with the increase in the number of scenes used for composition. While it is possible to composite multiple scenes when leveraging Neural-RF, rendering times are a strong limitation, specifically when employed in the case of XR applications. On the other hand, Explicit Lattice representations demand a large amount of memory, leading to infeasibility. On the other hand, our FusedRF representation alleviates the issues by fusing the

compositions iteratively, constraining memory, and rendering budgets to that of a single RF.

Quantitive Results: Along with maintaining tighter memory and rendering budgets, our proposed FusedRF representation also retains the quality of composited scenes. To validate this, we provide quantitive metrics of our FusedRF representation against the naive composition (Refer Tab. 3). This validates the representative capacity of our FusedRF representation.

Qualitative Results: The Qualitative results of our method are presented in the Figs. 1 and 3.

#	Neural Based		Voxel Based	
	NeRF	DVGO	TensoRF	Ours
1	10 M / 20s	800 M / 2.01s	25 M / 2.05s	25 M / 2.04s
2	20 M / 40s	1.6 G / 4.11s	50 M / 4.25s	25 M / 2.05s
4	40 M / 80s	3.2 G / 8.58s	100 M / 8.7s	25 M / 2.04s
8	80 M / 160s	OOM	200 M / 17.2s	25 M / 2.04s

Table 2. Rendering composition of RFs is *slow* and *memory intensive* as the number of scenes increases. Our proposed FusedRF performs fusion once, maintaining the memory and computing constant even when the number of composited scenes increases. (M: MB, G: GB, s: seconds). Experimented on RTX 3060 Ti (8GB).

Scene	Ours (w/o RGB)	Ours (Full)
Figure 1	36.71	39.89
Figure 3b	37.20	40.32
Figure 3d	35.18	38.78

Table 3. This table shows the PSNR of the images from some scenes in the paper. Please note that the PSNR reported is of the FusedNeRF images against composed NeRF images.

5. Conclusion

We present FusedRF, a method to create a single RF representation for a scene composed of multiple RFs. This reduces the memory and rendering overheads without degradation of quality. We showed our method over TensoRF [6] representation here. However, our method can be extended to any RF representation that uses explicit 3D lattices like InstantNGP, DVGO, Plenoxels, etc. [12, 25, 30]. As our method provides tighter memory and rendering budgets, using our FusedRF in XR applications like [9, 19] can facilitate the composition of multiple RFs while maintaining real-time results. The supplementary video provides an overview of our method, the multiview visualization, and the iterative addition of scenes in our results.

References

- [1] Sketchfab. <https://sketchfab.com/alban/models>. 1
- [2] Mallikarjun B R, Ayush Tewari, Xingang Pan, Mohamed Elgharib, and Christian Theobalt. gCoRF: Generative Compositional Radiance Fields. 2022. 1
- [3] Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1, 3, 4
- [4] Mark Boss, Raphael Braun, Varun Jampani, Jonathan T. Barron, Ce Liu, and Hendrik P.A. Lensch. NeRD: Neural Reflectance Decomposition from Image Collections. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. 2
- [5] Mark Boss, Varun Jampani, Raphael Braun, Ce Liu, Jonathan T. Barron, and Hendrik P.A. Lensch. Neural-PIL: Neural Pre-Integrated Lighting for Reflectance Decomposition. In *Adv. Neural Inform. Process. Syst.*, 2021. 2
- [6] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. TensorRF: Tensorial Radiance Fields. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022. 1, 2, 4
- [7] Pei-Ze Chiang, Meng-Shiun Tsai, Hung-Yu Tseng, Weisheng Lai, and Wei-Chen Chiu. Stylizing 3D Scene via Implicit Representation and HyperNetwork, 2021. 2
- [8] Chong Bao and Bangbang Yang, Zeng Junyi, Bao Hujun, Zhang Yinda, Cui Zhaopeng, and Zhang Guofeng. NeuMesh: Learning Disentangled Neural Mesh-based Implicit Field for Geometry and Texture Editing. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022. 1, 2
- [9] Nianchen Deng, Zhenyi He, Jiannan Ye, Budmonde Duinkharjav, Praneeth Chakravarthula, Xubo Yang, and Qi Sun. FoV-NeRF: Foveated Neural Radiance Fields for Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*, 2022. 4
- [10] Danny Driess, Zhiao Huang, Yunzhu Li, Russ Tedrake, and Marc Toussaint. Learning multi-object dynamics with compositional neural radiance fields. In *6th Annual Conference on Robot Learning*, 2022. 1
- [11] Shuangkang Fang, Weixin Xu, Heng Wang, Yi Yang, Yufeng Wang, and Shuchang Zhou. One is all: Bridging the gap between neural radiance fields architectures with progressive volume distillation. *AAAI*, 2023. 2
- [12] Fridovich-Keil and Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance Fields without Neural Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 2, 4
- [13] Rahul Goel, Dhawal Sirikonda, Saurabh Saini, and P.J. Narayanan. Interactive Segmentation of Radiance Fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 2, 3
- [14] Yi-Hua Huang, Yue He, Yu-Jie Yuan, Yu-Kun Lai, and Lin Gao. StylizedNeRF: Consistent 3D Scene Stylization as Stylized NeRF via 2D-3D Mutual Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 2
- [15] Clément Jambon, Bernhard Kerbl, Georgios Kopanas, Stavros Diolatzis, Thomas Leimkühler, and George Dretakis. NeRFshop: Interactive Editing of Neural Radiance Fields. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 2023. 1, 2
- [16] Wei Jiang, Kwang Moo Yi, Golnoosh Samei, Oncel Tuzel, and Anurag Ranjan. NeuMan: Neural Human Radiance Field from a Single Video. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022. 1
- [17] Sosuke Kobayashi, Eiichi Matsumoto, and Vincent Sitzmann. Decomposing NeRF for Editing via Feature Field Distillation. In *Adv. Neural Inform. Process. Syst.*, 2022. 2
- [18] Verica Lazova, Vladimir Guzov, Kyle Olszewski, Sergey Tulyakov, and Gerard Pons-Moll. Control-NeRF: Editable Feature Volumes for Scene Rendering and Manipulation. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2023. 1, 2
- [19] Chaojian Li, Sixu Li, Yang Zhao, Wenbo Zhu, and Yingyan Lin. RT-NeRF: Real-Time On-Device Neural Radiance Fields Towards Immersive AR/VR Rendering. In *Proceedings of the 41st IEEE/ACM International Conference on Computer-Aided Design*, 2022. 4
- [20] Shaoxu Li and Ye Pan. Interactive Geometry Editing of Neural Radiance Fields, 2023. 1, 2
- [21] Ruofan Liang, Jiahao Zhang, Haoda Li, Chen Yang, Yushi Guan, and Nandita Vijaykumar. SPIDR: SDF-based Neural Point Fields for Illumination and Deformation. *arXiv preprint arXiv:2210.08398*, 2022. 1
- [22] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020. 1, 2, 3, 4
- [23] Ashkan Mirzaei, Tristan Aumentado-Armstrong, Konstantinos G. Derpanis, Jonathan Kelly, Marcus A. Brubaker, Igor Gilitschenski, and Alex Levinshtein. SPIn-NeRF: Multiview Segmentation and Perceptual Inpainting with Neural Radiance Fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 1
- [24] Ashkan Mirzaei, Tristan Aumentado-Armstrong, Konstantinos G. Derpanis, Jonathan Kelly, Marcus A. Brubaker, Igor Gilitschenski, and Alex Levinshtein. SPIn-NeRF: Multiview segmentation and perceptual inpainting with neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 2
- [25] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Trans. Graph.*, 2022. 2, 4
- [26] Jacob Munkberg, Jon Hasselgren, Tianchang Shen, Jun Gao, Wenzheng Chen, Alex Evans, Thomas Mueller, and Sanja Fidler. Extracting Triangular 3D Models, Materials, and Lighting From Images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 2

- [27] Thu Nguyen-Phuoc, Feng Liu, and Lei Xiao. SNeRF: Stylized Neural Implicit Representations for 3D Scenes. *ACM Trans. Graph.*, 2022. 2
- [28] Yi-Ling Qiao, Alexander Gao, and Ming Lin. NeuPhysics: Editable Neural Geometry and Physics from Monocular Videos. In *Adv. Neural Inform. Process. Syst.*, 2022. 1
- [29] Pratul P. Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T. Barron. NeRV: Neural Reflectance and Visibility Fields for Relighting and View Synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 2
- [30] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct Voxel Grid Optimization: Super-fast Convergence for Radiance Fields Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 1, 2, 3, 4
- [31] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Improved Direct Voxel Grid Optimization for Radiance Fields Reconstruction. *arXiv,abs/2206.05085*, 2022. 1, 4
- [32] Andrea Tagliasacchi and Ben Mildenhall. Volume Rendering Digest (for NeRF). *Technical Report*, 2022. 3
- [33] Konstantinos Tertikas, Despoina Paschalidou, Boxiao Pan, Jeong Joon Park, Mikaela Angelina Uy, Ioannis Emiris, Yanis Avrithis, and Leonidas Guibas. PartNeRF: Generating Part-Aware Editable 3D Shapes without 3D Supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 1
- [34] Vadim Tschernezki, Iro Laina, Diane Larlus, and Andrea Vedaldi. Neural Feature Fusion Fields: 3D distillation of self-supervised 2D image representations. In *International Conference on 3D Vision (3DV)*, 2022. 2
- [35] Tianyu Wang, miaomiao Liu, and Kee Siong Ng. Scalable 3D Object-centric Learning. 2023. 1, 2
- [36] Silvan Weder, Guillermo Garcia-Hernando, Áron Monszpart, Marc Pollefeys, Gabriel Brostow, Michael Firman, and Sara Vicente. Removing Objects From Neural Radiance Fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 1
- [37] Qianyi Wu, Xian Liu, Yuedong Chen, Kejie Li, Chuanxia Zheng, Jianfei Cai, and Jianmin Zheng. Object-Compositional Neural Implicit Surfaces. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022. 1
- [38] Tianhao Wu, Fangcheng Zhong, Andrea Tagliasacchi, Forrester Cole, and Cengiz Oztireli. D²NeRF: Self-Supervised Decoupling of Dynamic and Static Objects from a Monocular Video. In *Adv. Neural Inform. Process. Syst.*, 2022. 1, 2
- [39] Tianhan Xu and Tatsuya Harada. Deforming Radiance Fields with Cages. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2022. 1, 2
- [40] Yinghao Xu, Menglei Chai, Zifan Shi, Sida Peng, Skokhodov Ivan, Siarohin Aliaksandr, Ceyuan Yang, Yujun Shen, Hsin-Ying Lee, Bolei Zhou, and Tulyakov Sergiy. DiscoScene: Spatially Disentangled Generative Radiance Field for Controllable 3D-aware Scene Synthesis. *arxiv:2212.11984*, 2022. 1, 2
- [41] Bangbang Yang, Yinda Zhang, Yinghao Xu, Yijin Li, Han Zhou, Hujun Bao, Guofeng Zhang, and Zhaopeng Cui. Learning Object-Compositional Neural Radiance Field for Editable Scene Rendering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. 1, 2
- [42] Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. PlenOctrees for Real-time Rendering of Neural Radiance Fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 1
- [43] Heng Yu, Koichiro Niinuma, and Laszlo A. Jeni. CoNFies: Controllable Neural Face Avatars. In *International Conference on Automatic Face and Gesture Recognition*, 2023. 1
- [44] Xiuming Zhang, Pratul P. Srinivasan, Boyang Deng, Paul Debevec, William T. Freeman, and Jonathan T. Barron. NeRFactor: Neural Factorization of Shape and Reflectance under an Unknown Illumination. *ACM Trans. Graph.*, 2021. 2
- [45] Chengwei Zheng, Wenbin Lin, and Feng Xu. EditableNeRF: Editing Topologically Varying Neural Radiance Fields by Key Points. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 1
- [46] Shuaifeng Zhi, Tristan Laidlow, Stefan Leutenegger, and Andrew Davison. In-Place Scene Labelling and Understanding with Implicit Scene Representation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. 2